

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (currently amended) A computer-implemented method of identifying candidate genes from a plurality of DNA sequences, the method comprising:
obtaining gene expression profile data for a plurality of DNA sequences, wherein the gene expression profile data describe behavioral patterns of gene expression;
identifying a group of DNA sequences for further analysis;
using information extraction algorithms to retrieve and extract pathway information from a database related to the group of DNA sequences;
cross-referencing said pathway information; and to said DNA sequences;
ranking the pathway information based on a ranking of a publication in a citation index;
viewing said cross-referenced information and said ranking; and,
wherein viewing the cross-referenced information and said ranking facilitates the identification of candidate genes.
2. (original) The computer-implemented method of Claim 1, wherein the pathway information is stored in a database.
3. (original) The computer-implemented method of Claim 2, wherein the cross-referenced information is stored in a database.
4. (original) The computer-implemented method of Claim 1, wherein the cross-referenced information is viewed as a directed graph.
5. (original) The computer-implemented method of Claim 1, wherein identifying a group of DNA sequences further comprises clustering the gene expression profile data to form clusters.
6. (original) The computer-implemented method of Claim 5, wherein clustering is unsupervised clustering.

7. (original) The computer-implemented method of Claim 5, wherein clustering is supervised clustering.

8. (original) The computer-implemented method of Claim 5, wherein clustering is a combination of supervised and unsupervised clustering.

9. (original) The computer-implemented method of Claim 5, wherein the group of DNA sequences represents a cluster.

10. (original) The computer-implemented method of Claim 1, wherein the gene expression profile data is derived from microarray experiments.

11. (original) The computer-implemented method of Claim 1, wherein the information extraction is performed using natural language processing algorithms.

12. (original) The computer-implemented method of Claim 11, wherein the natural language processing algorithms include template filling or Hidden Markov-Models.

13. (original) The computer-implemented method of Claim 11, wherein an information extraction algorithm utilizes a text comparison algorithm.

14. (original) The computer-implemented method of Claim 1, wherein the information is extracted from one or more literature databases from the group consisting of MEDLINE, USPTO patent published patent database, USPTO issued patent database, the WIPO patent database, and the KEGG, MIPS and OMIM database.

15. (canceled)

16. (currently amended) A data processing system for identifying candidate genes from a plurality of DNA sequences of known expression pattern, comprising:

a processor; and,

a memory coupled to the processor, wherein the memory has ~~configured to store~~ instructions for execution by the processor, the instructions comprising:

instructions for accessing and extracting pathway information from a literature

database comprising a biomedical publication;

instructions for cross-referencing said pathway information to said candidate genes;

instructions for ranking the biomedical publication and instructions to assign a ranking score to the pathway information extracted from a biomedical publication based on the ranking of the biomedical publication; and,

instructions for viewing said cross-referenced information and said ranking score.

17. (currently amended) The data processing system of Claim 16, wherein said executable instructions further comprise instructions for storing said pathway information and said cross-referenced information in a database.

18. (canceled)

19. (canceled)

20. (currently amended) A data processing system for identifying candidate genes from a plurality of DNA sequences, comprising:

a processor; and,

a memory coupled to the processor, wherein the memory has ~~configured to store~~ instructions for execution by the processor, the instructions comprising:

instructions for clustering the plurality of DNA sequences based on the behavioral patterns of the DNA sequences as described by gene expression profile data;

instructions for accessing and extracting pathway information from a literature database comprising a biomedical publication;

instructions for cross-referencing said pathway information to said candidate genes;

instructions for ranking the biomedical publication and instructions to assign a ranking score to the pathway information extracted from a biomedical publication based on the ranking of the biomedical publication; and,

instructions for viewing said cross-referenced information and said ranking score.